Spectral Line Data Reduction

- Joe Callingham

- Kenya 2018
- Based on talks by Katharine Johnston (University of Leeds) and Luke Hindson

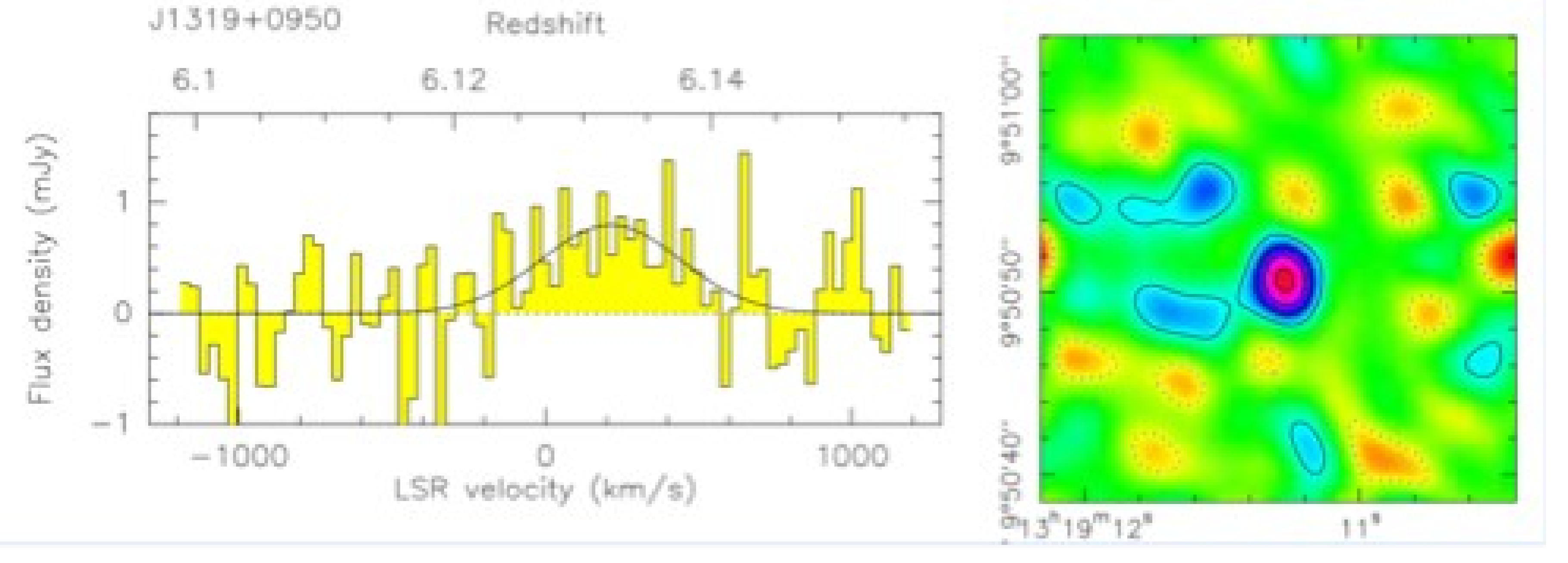
Introduction

- We have discussed some of the science you can do with spectral lines
- In this lecture we will explore the important differences between spectral line and continuum data reduction
  - Setting up your observations
  - Data reduction
  - Image and spectral analysis

What is spectral line interferometry

- Observing many adjacent frequency channels with an interferometer for an object whose flux changes rapidly with frequency

- Results in a third axis: frequency or velocity

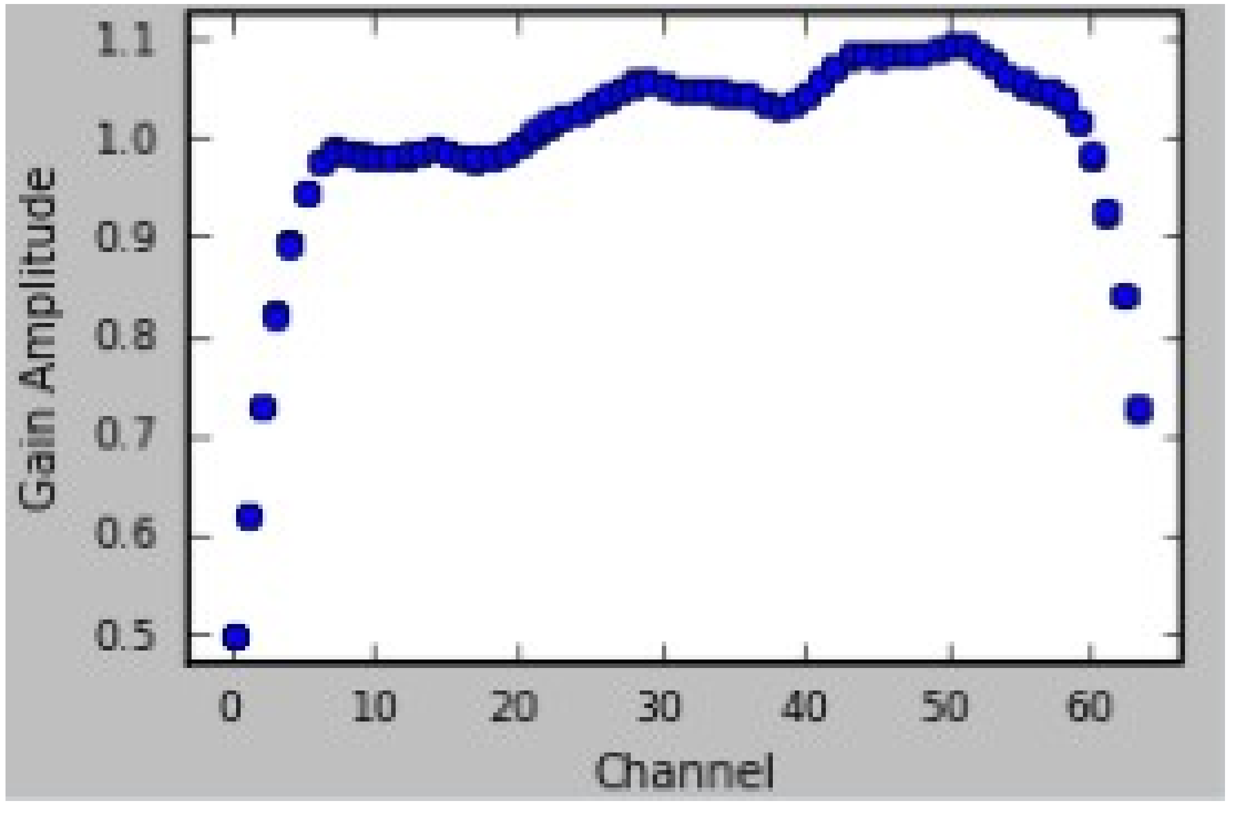


CO (6-5) emission from a high redshift quasar observed with the Plateau de Bure interferometer (Wang et al. 2010)

What is spectral line interferometry

- However, today even continuum observations are carried out with many channels (called pseudo continuum)

Plot of response across a band for one VLA antenna as a function of channel during the reduction of continuum data



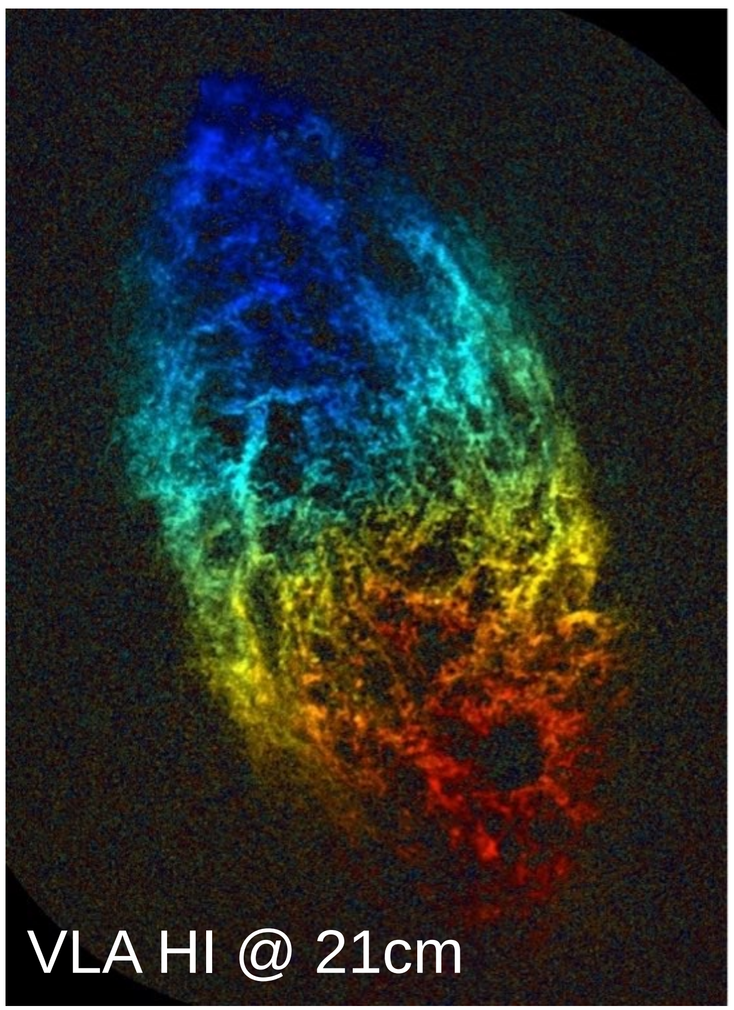
- Therefore, you need to understand spectral line interferometry to do (almost all) interferometry

Setting up you observations

1. Choose your science case and lines(s)
2. choose your source(s) and research its properties
3. Choose your interferometer and configuration
4. Check source velocity reference frame, on-line Doppler tracking
5. Choose the channel width, total bandwidth, and number of channels
6. Determine your required sensitivity and time on source
7. Choose your calibrators
8. Data rate and size considerations

1. Choose your science case and line(s)

Doppler shifts → Velocity fields



Velocity field of galaxy M33 in HI line. Colours show Doppler shift of line and brightness is proportional to HI column density (NRAO, Thilker et al.

1. Choose spectral line(s) for science case
2. Find which telescopes have bands that can observe the line(s)

Physical properties from lines for science case

Column density

Chemistry

Excitation Temperature

Optical depth

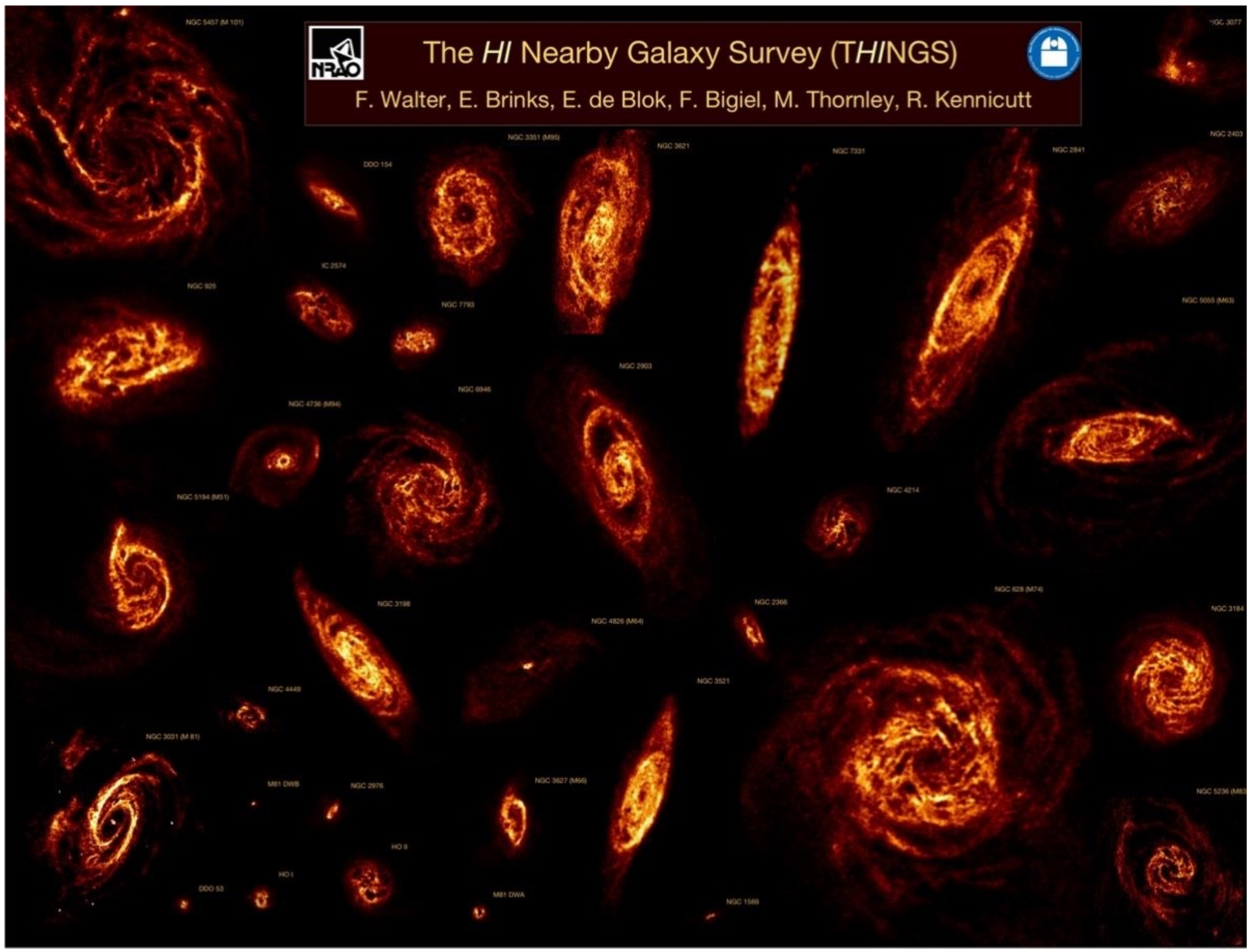
Dynamics

Magnetic field strength

Turbulent motions

Density

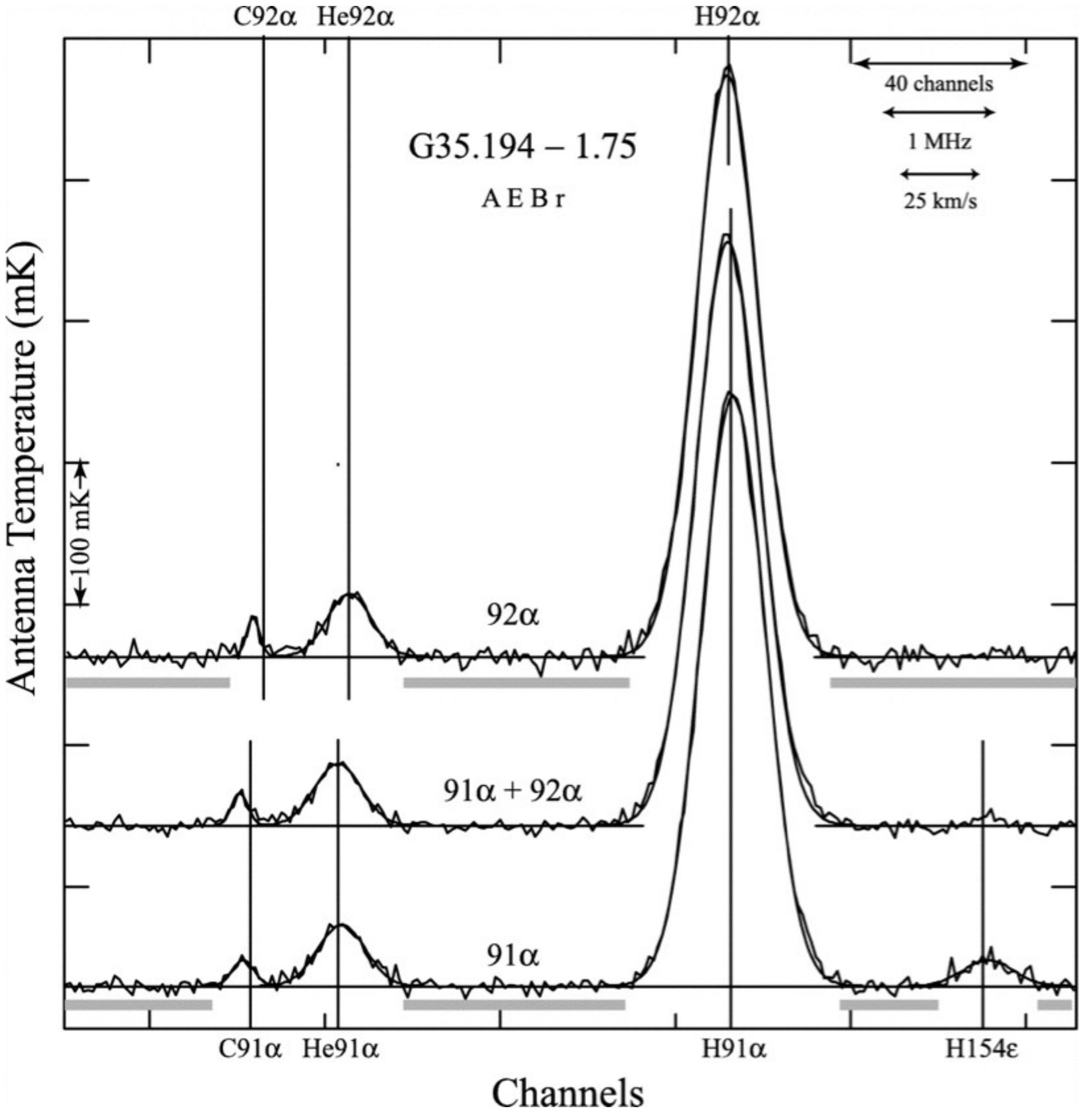
Commonly observed lines: HI line at 21cm (1.4 GHz)



Can determine:

- Morphology of atomic gas
- Dynamics
  - gives enclose mass (inc dark matter)
- column density and mass of optically thin gas
- temperature of optically thick gas
- distance via Hubble law

Commonly observed lines: Radio Recombination Lines (RRLs)



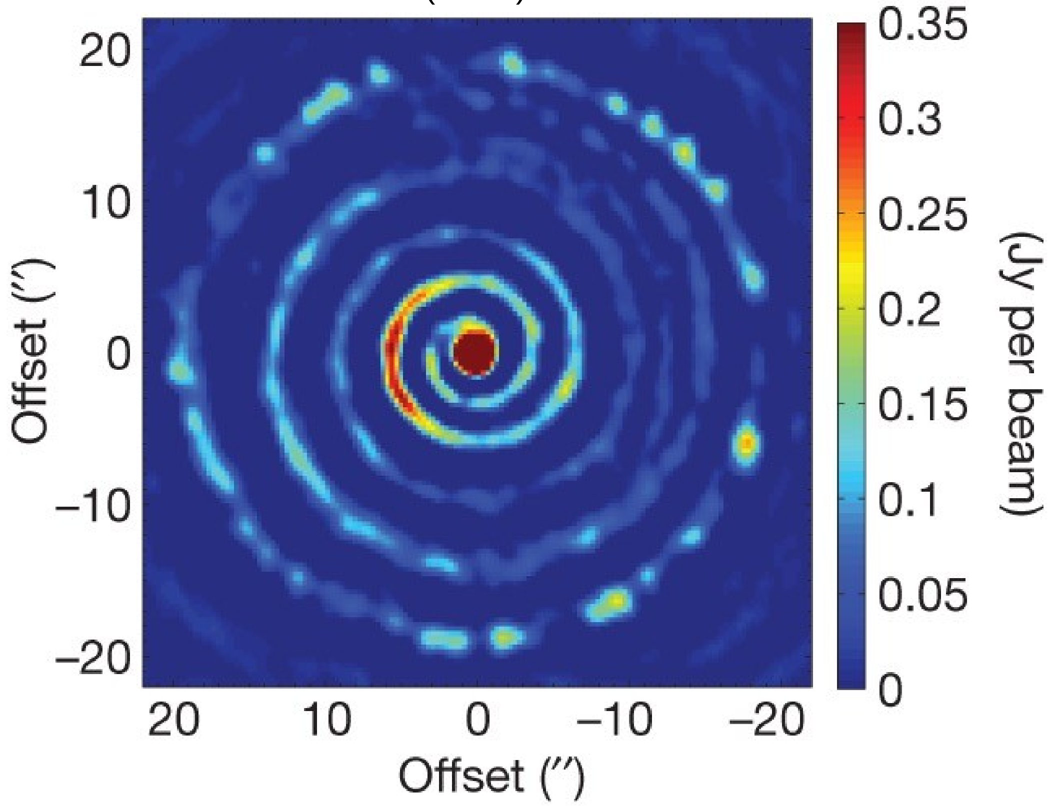
RRLs for n=91 and 92 for H, He and C in an HII region (Quireza et al. 2006)

Can determine:

- Ionised gas dynamics
- optical depth and electron temperature (in Local Thermodynamic Equilibrium, LTE)
- Gas density from collisional broadening of the lines
- Speed of turbulent motions from linewidths (for low n)
- Abundance ratio of H/He/C
- Magnetic field strengths from Zeeman splitting

Commonly observed lines: Molecular lines

AGB star R Sculptors
CO (J=3-2) with ALMA



Maercker et al. (2012)

Common molecular lines:

- CO, OH, H2O, SiO, HCN, methanol, NH3

Can also observe absorption lines and masers.

Can determine:

- Morphology of molecular gas
- Dynamics
- Gas column and volume density
- Optical depth
- Temperature
- Abundances
- Magnetic field strength
- Chemistry
- Presence of shocks (SiO)

2. Choose your source and research its properties

Learn about the object you wish to study e.g

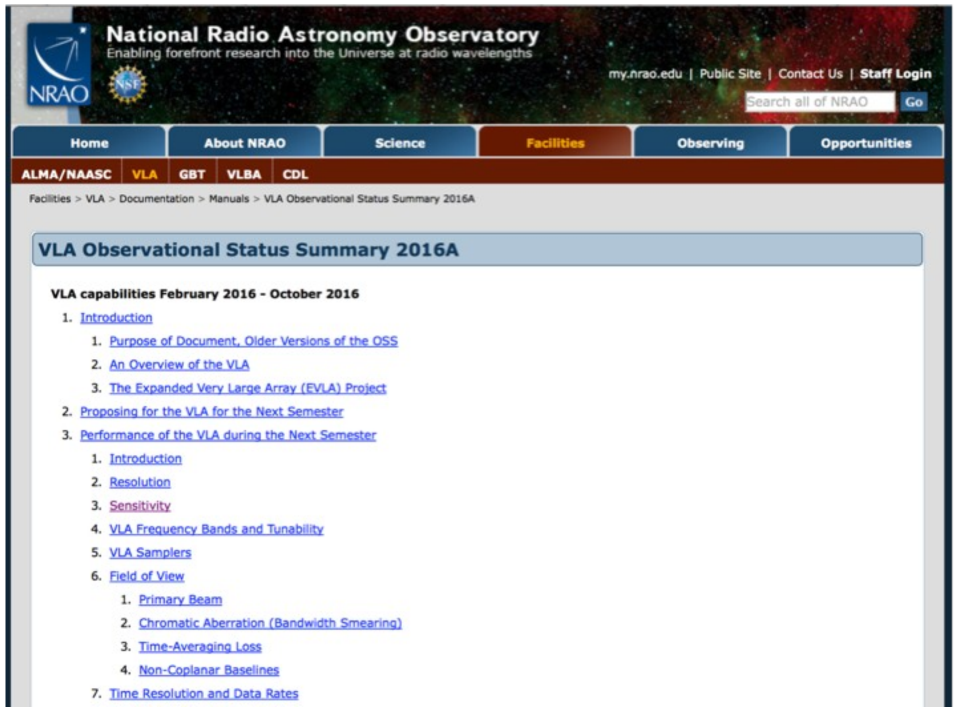
- Position (and equinox)
- Size
- Velocity (and reference frame)
- Estimated brightness

Search for literature on your source using resources like:

- ADS (http://www.adsabs.harvard.edu/)
- NED (https://ned.ipac.caltech.edu/)
- CDS (http://cds.u-strasbg.fr/)

Choose your interferometer and configuration

- Large number of possible instruments to name just a few:
  - (J)VLA, ALMA, MWA, LOFAR, PdB, ATCA, EVN, AVN
- Read about the telescope online e.g for the VLA



3. Choose your interferometer and configuration(s)

(J)VLA

Larger arrays: better resolution! Good for absorption studies as want to detect a small, bright background source

Part of the ALMA Compact Array

Smaller arrays: have better surface brightness sensitivity so easier to detect extended or faint emission

4. Check source velocity reference frame

Need to always specify the velocity reference frame

Rest Frame	Corrected for	Amplitude of Correction (km/s)
Topocentric	Nothing	0
Geocentric	Earth rotation	0.5
Earth-Moon Barycentric	Effect of Moon on Earth	0.013
Heliocentric	Earth's orbital motion	30
Solar System Barycentric	Effect of planets on Sun	0.012
Local Standard of Rest (LSRK/D)	Solar motion	20
Galactocentric	Milky Way Rotation	230
Local Group Barycentric	Milky Way Motion	~100
Virgocentric	Local Group Motion	~300
Microwave background	Local Supercluster motion	~600

4. Check source velocity reference frame

- Optical Barycentric or Heliocentric system often used for extragalactic observations

- Radio Local Standard of Rest (LSR) used for Milky Way observations

$$\frac{v_{\rm radio}}{c} = \frac{\nu_0 - \nu}{\nu_0}$$

Doppler shift and Doppler Tracking

- **on-line Doppler tracking** automatically corrects to a given reference frame during the observation in real time
- The tracked or observed frequency is usually called the **sky frequency**
- However, for wide frequency bands (most modern interferometers) on-line Doppler tracking is not done / recommended as correction is only strictly correct at one frequency
- Instead **Doppler Setting** is used, i.e. sky frequency calculated once at the start of the observation
- Further corrections can be made during the data reduction and imaging stages (need > 4 channels across the line for good correction)

5. Choose the channel width, total bandwidth and number of channels

- Channel width determined by required **spectral resolution** and **sensitivity**

- Total bandwidth should:

  - Leave good line free channels at the ends of the band for continuum subtraction (end channels often bad)

- In a “lag” correlation, total number of channels is conserved, so total bandwidth is directly related to channel width

Velocity Coverage

- For radio/millimetre frequency observations:

$$\frac{\Delta f}{f} \simeq \frac{\Delta v}{c}$$

Where *Δf* is the bandwidth (Hz), *f* the rest frequency (Hz), *Δv* is the velocity span ($\mathrm{kms^{-1}}$) and *c* = 3 x 10^5 $\mathrm{kms^{-1}}$, the speed of light

- So if you want a velocity coverage of 100 $\mathrm{kms^{-1}}$ at a frequency of 22.2 GHz then:

$$\Delta f = \frac{100}{3 \times 10^5} \times 22.2 \times 10^9 = 7.4\,\mathrm{MHz}$$

- Allow some leeway for filter edge effects.

Velocity Resolution

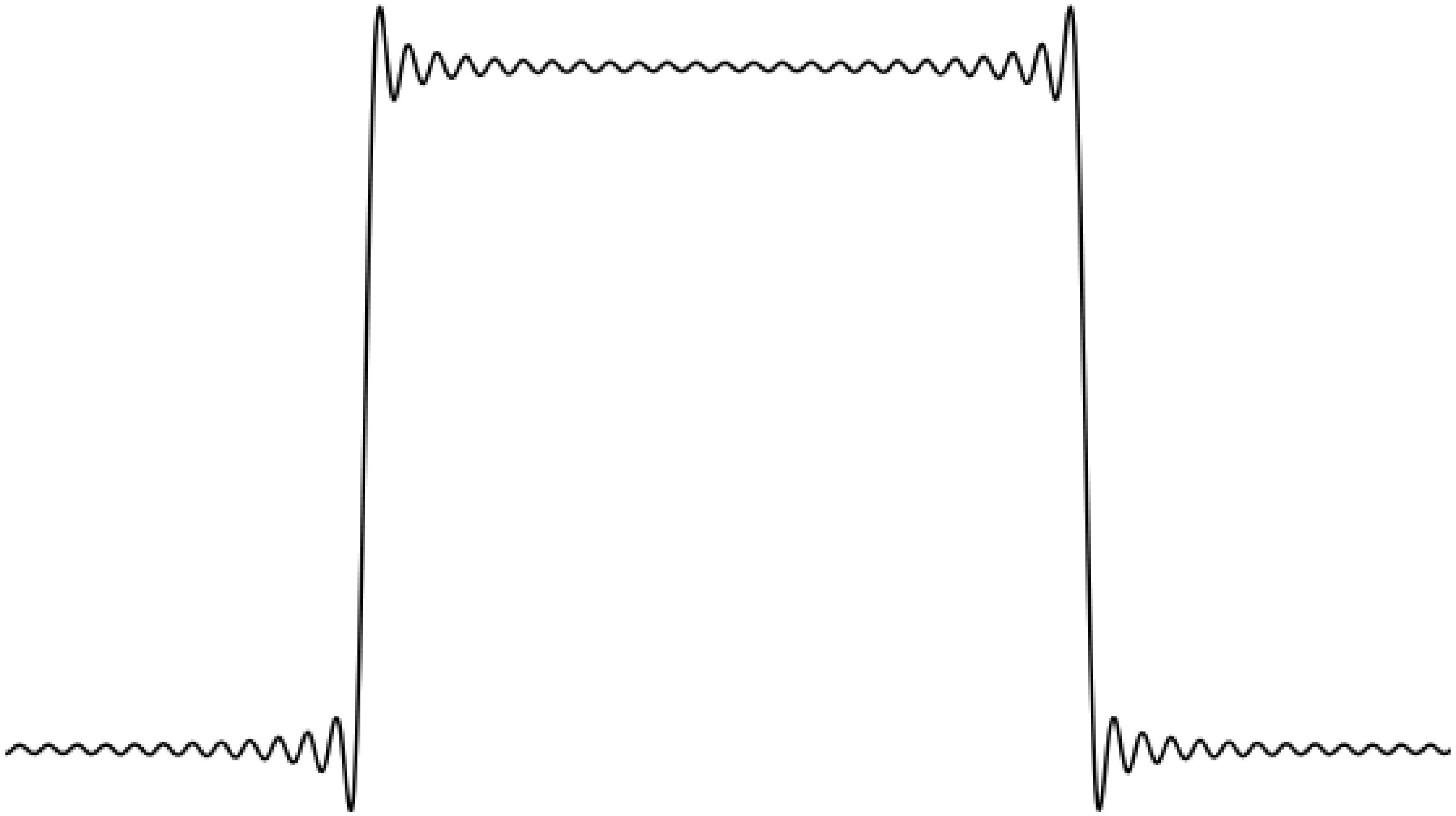
- It is good practice to have a velocity resolution with at least 3 channels between the half power points of a typical spectral line
- So if you have lines with typical FWHM of 1 km/s at 22.2 GHz then you want a spectral resolution of:

$$\Delta f = \frac{1/3}{3 \times 10^5} \times 22.2 \times 10^9 = 25\,\mathrm{kHz}$$

- Many correlates have a maximum channel bandwidth product. Often you want to select the configuration with maximum bandwidth for the required velocity resolution

Gibbs Ringing and Hanning Smoothing

Gibbs Ringing



- Seen at channel edges
- Also see for bright lines e.g masers, RFI

Fixed by

Hanning Smoothing

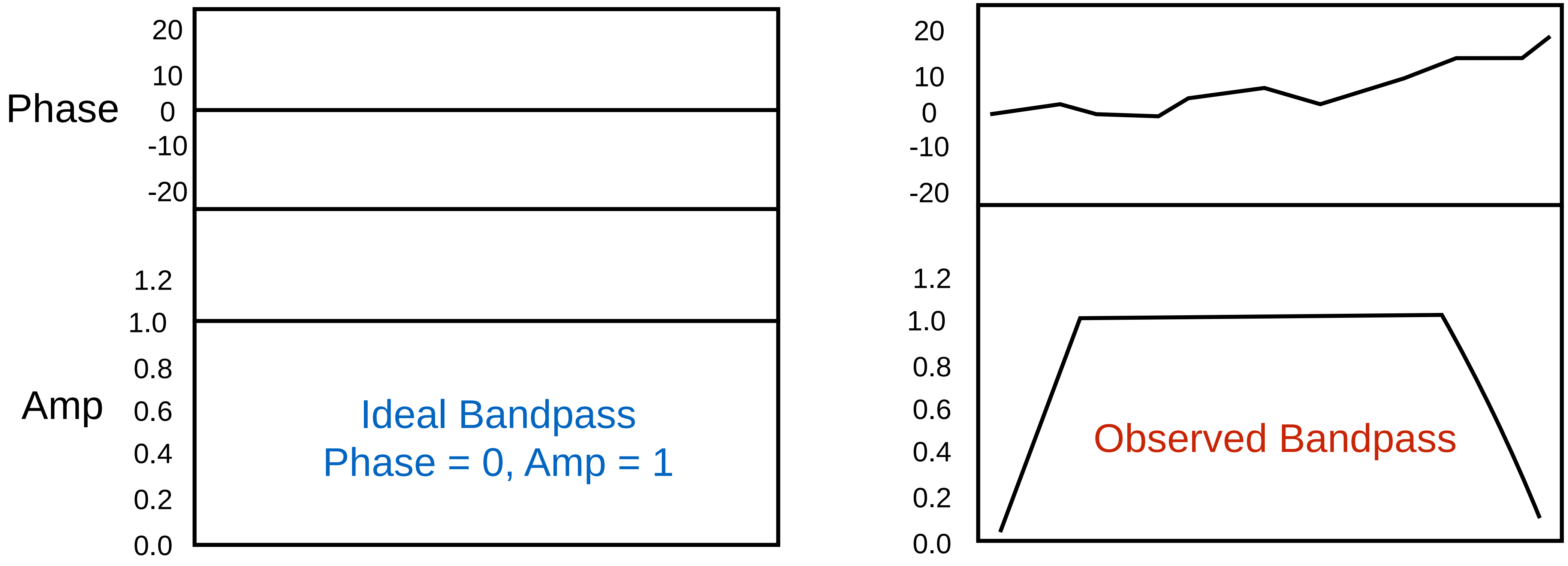
- Smooth data with a running mean in frequency with weightsL
  - 0.5 for central channels
  - 0.25 for adjacent channels
  - i.e a triangular smoothing kernel
- **Pro:** reduce ringing
- **Con:** reduces spectral resolution by 2

6. Determine your required channel sensitivity and time on source

- Determine how many channels needed to adequately resolve your line (e.g. >5)
- If **detection is important**, need to detect line **peak channel** with >3-5 sigma
- If **need to resolve emission in different channels** for e.g. dynamics, need to detect line in **faintest of these channels** with >3-5 sigma
- You should determine the estimated flux of the source (May need to convert from brightness temperature T_b or expected column density)
- Use sensitivity calculator to determine required time on-source (e.g. VLA, ALMA, SMA, ATCA)

7. Choose your calibrators
Bandpass calibration

The bandpass is the spectral frequency response of an antenna to spectrally flat source of unit amplitude



The observed phase and amplitude is not ideal because of delay error and limitation of the antenna transmission and electronics

Why is bandpass calibration important?

The quality of the bandpass calibration is a key limiting factor in the ability to detect and analyse spectral features

v-dependant amplitude errors:

- limit ability to detect/measure weak emission/absorption lines superposed on the continuum
- may mimic changes in structure

v-dependant phase errors:

- may lead to spurious positional offsets between spectral features as a function of frequency, imitating Doppler motions
  - Relative positional accuracy in channel images $\Delta\theta/\theta_B = \Delta\phi/360^\circ$ where θ_B is the synthesised beam and $\Delta\phi$ is the scatter in the phase

For pseudo-continuum experiments conducted in spectral line mode, dynamic range of final images is limited by bandpass quality

What makes a good bandpass calibrator?

Select a bright continuum source with:

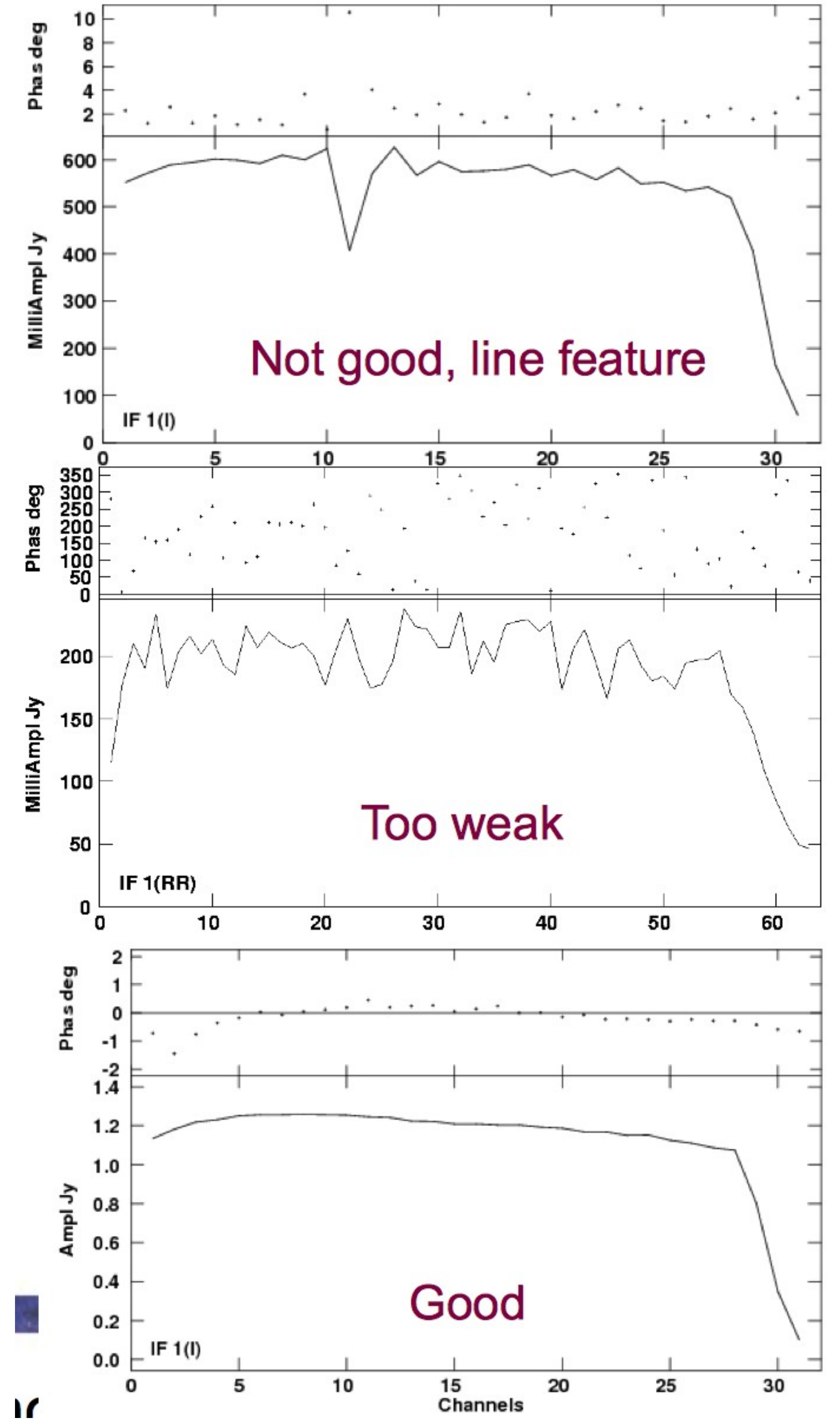
- High SNR in each channel
- Intrinsically flat spectrum
- No spectral lines/features
- No changes in structure across band (e.g. point source at all frequencies

Calibration should not contribute to noise in target spectrum, i.e. in one channel:

bandpass SNR > target SNR

Can smooth the bandpass or fit polynomial to increase the SNR

May need a higher SNR for bandpass calibrator if looking for faint lines on strong continuum

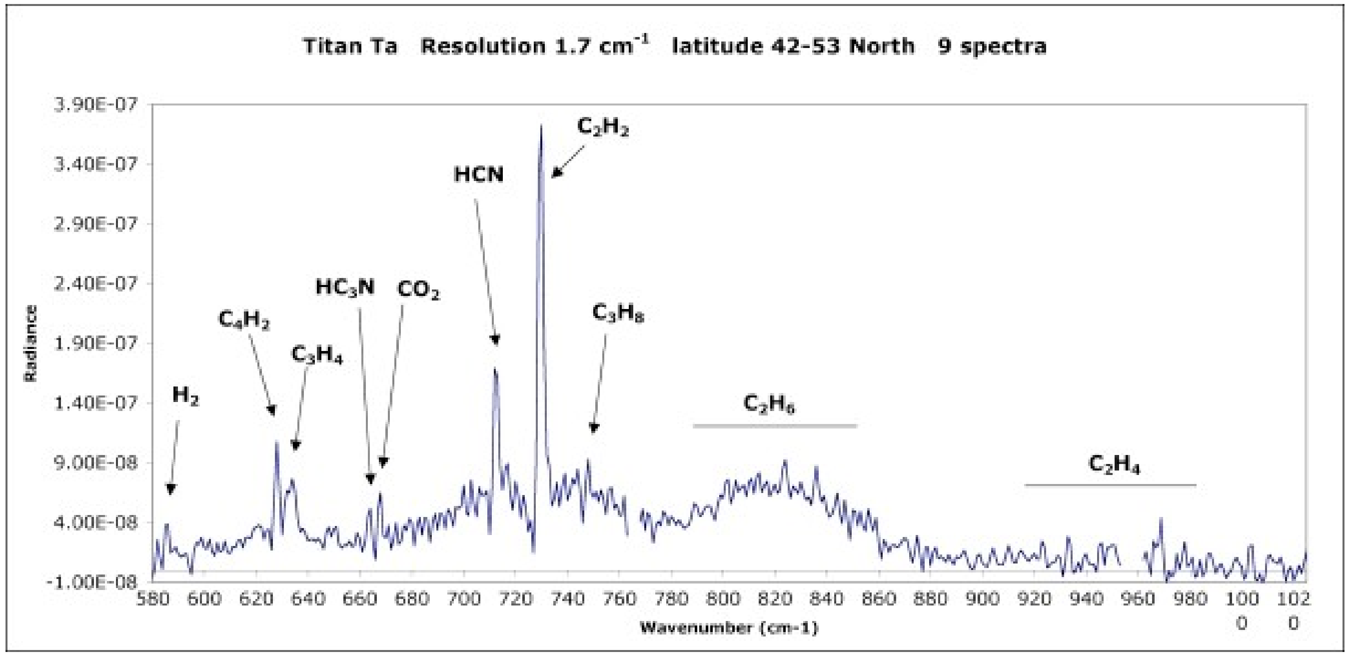


7. Choose your calibrators
Flux calibration

- In mm/sub-mm observing, if using solar system object with an atmosphere for calibration (Jovian or Saturnian moons, for example), be aware that these objects often have absorption lines

- Check your calibrator spectrum carefully, particularly if that source is not red-shifted

- You should exclude affected channels (or even baseboards if the absorption is broad)



Titan clearly shows multiple emission features

8. Data rate and size considerations

Finally, many observatories impose a maximum data rate for a given observation

For instance, for the VLA the data rate depends on:

- The number of channels
- The number of spectral windows
- The number of polarisations observed
- The length of one integration in seconds
- the number of antennas

It is not uncommon for datasets to be in the many 100s of GBs. Large datasets are difficult to transfer and reduce…. only use high spectral and time resolution if you really need it.

Data reduction

- Flagging methods
- Bandpass calibration
- Doppler correction
- Continuum subtraction
- Self-calibration
- Imaging of cubes

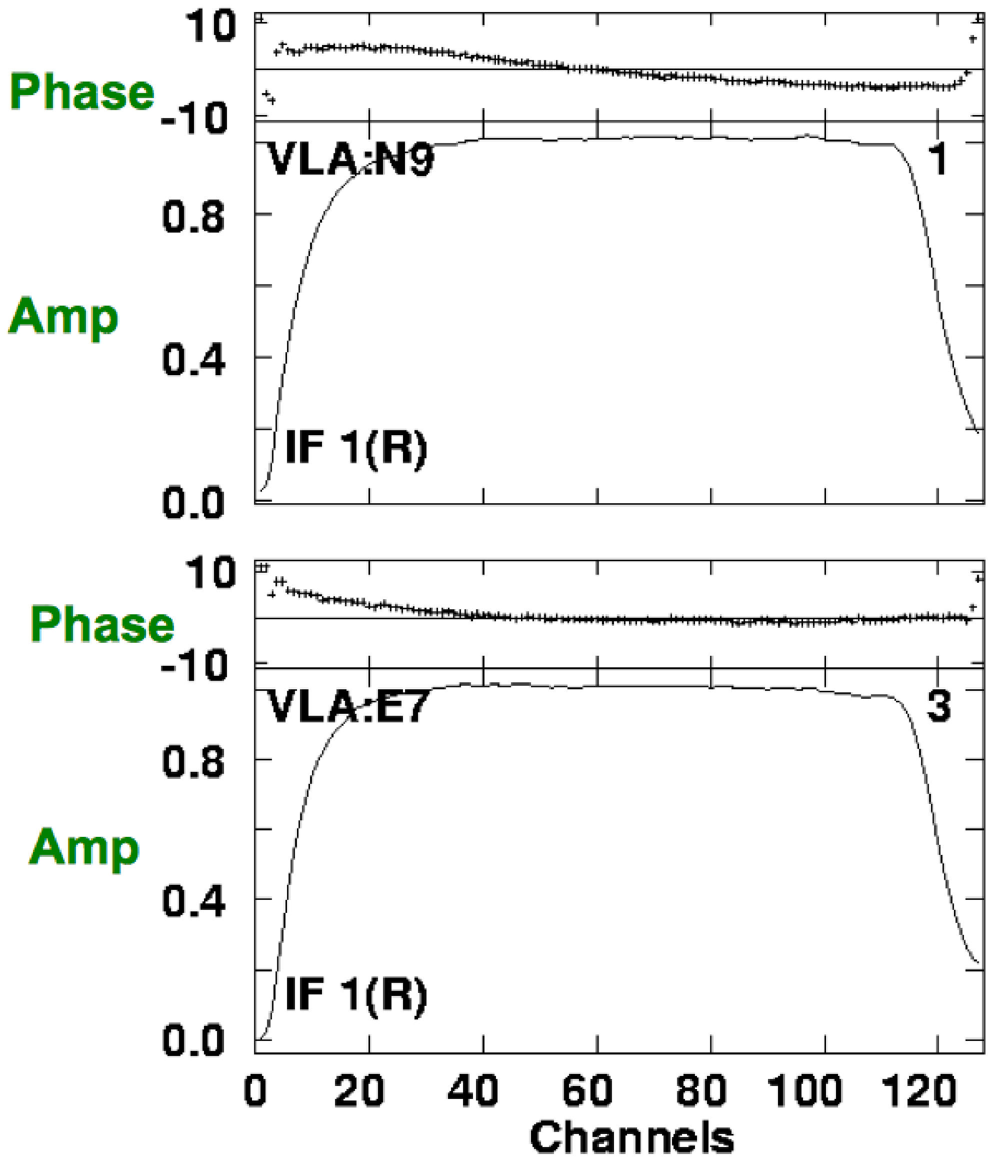
Flaggin Methods - in frequency

- For large data sets, checking the data channel-by channel is not practical
- This task can be simplified using approaches such as:
  - Examination of scalar-averaged cross-power spectra: check for dips or spikes
  - Use of automated flagging routines: these can flag based on deviation from expected spectral behaviour (e.g. SERPent, for e-MERLIN, AOFlagger)

But… if you are planning to make image cubes then avoid excessive frequency dependant flagging which changes the *uv*-coverage across the band

Bandpass calibration

Has your bandpass calibration gone well?



Examples of good-quality bandpass solutions for 2 antenna

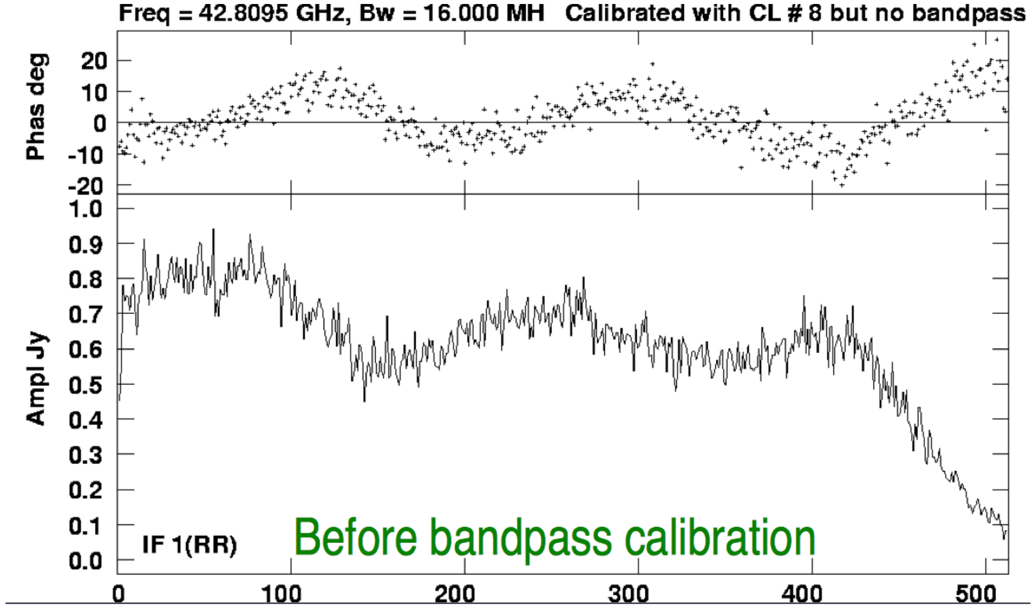
- Solutions should look comparable for all antennas
- Mean amplitude ~1 across useable portion of the band
- No sharp variations in the amplitude and phase
- Variations are not dominated by noise

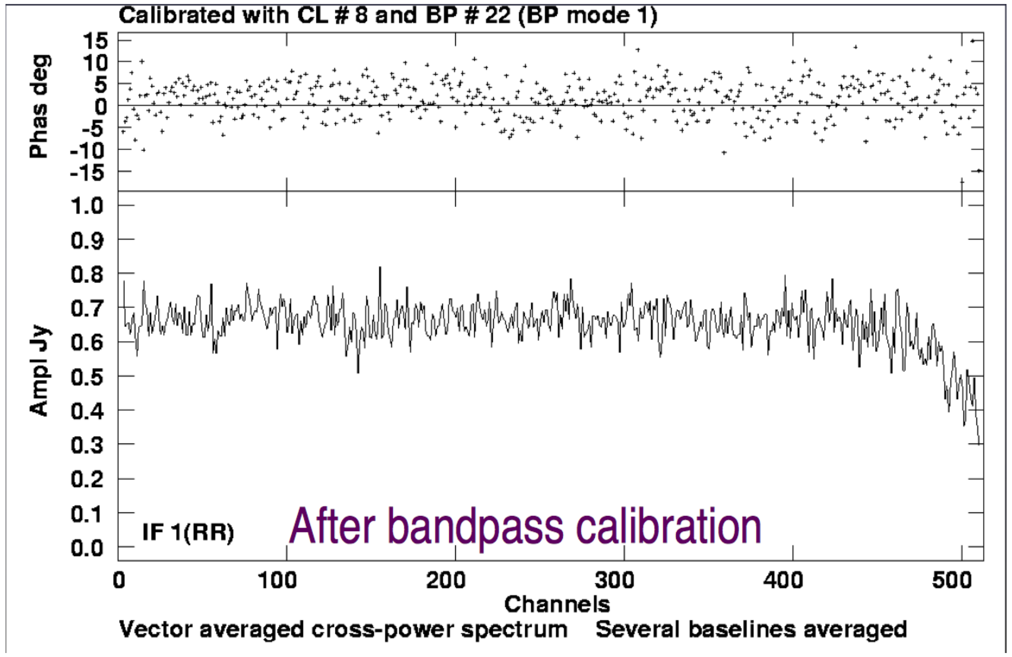
Bandpass calibration

Another good way to check: examine cross-power spectra of a continuum (flat spectrum) source with BP corrections applied

Checklist:

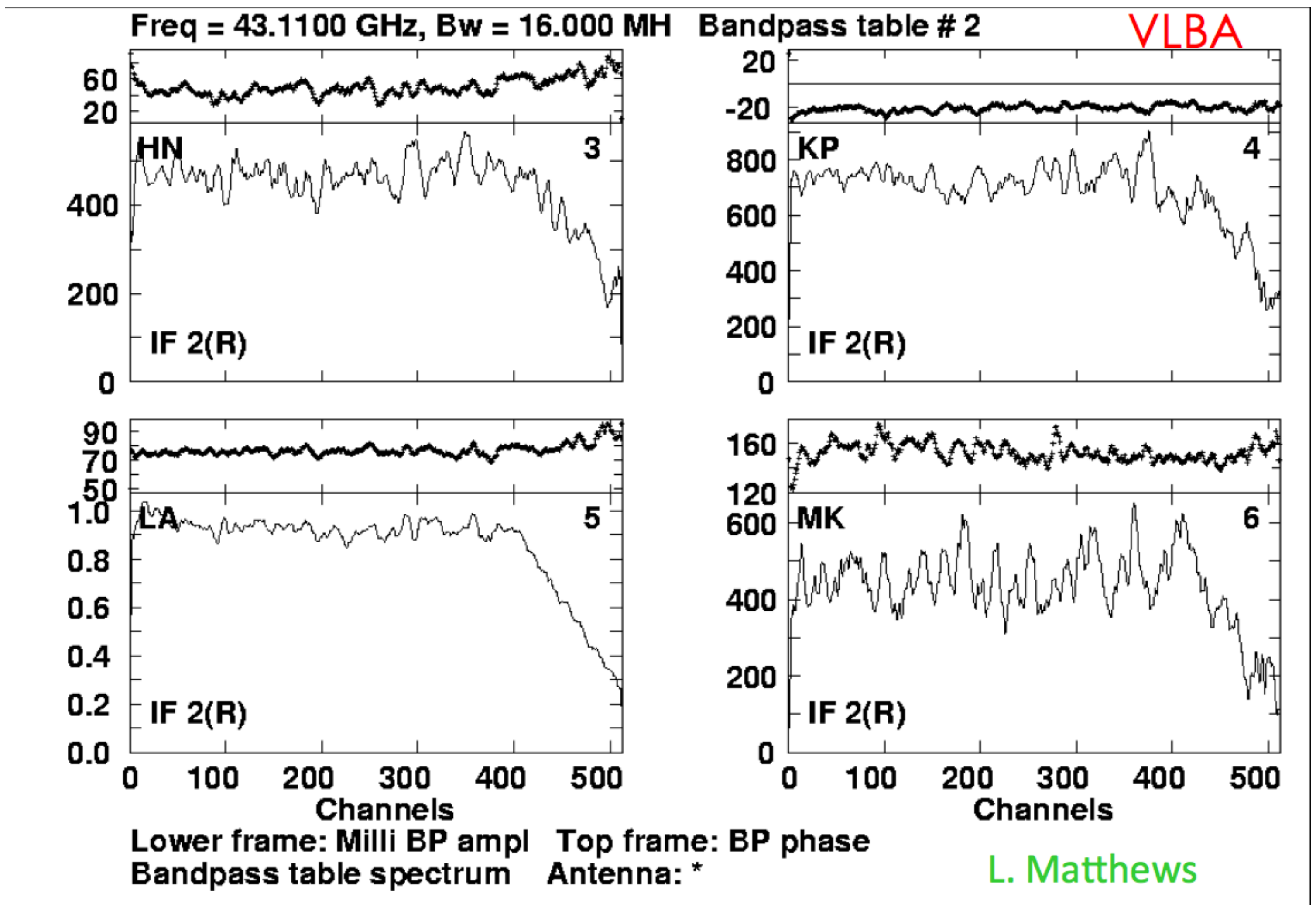
- ✓Phases are flat across the band
- ✓Amplitude is constant across the band
- ✓Corrected data do not have significantly increased noise
- ✓absolute flux level is not biased high or low





Bandpass calibration

When has your bandpass calibration gone **bad?**



The solutions shown are clearly bad

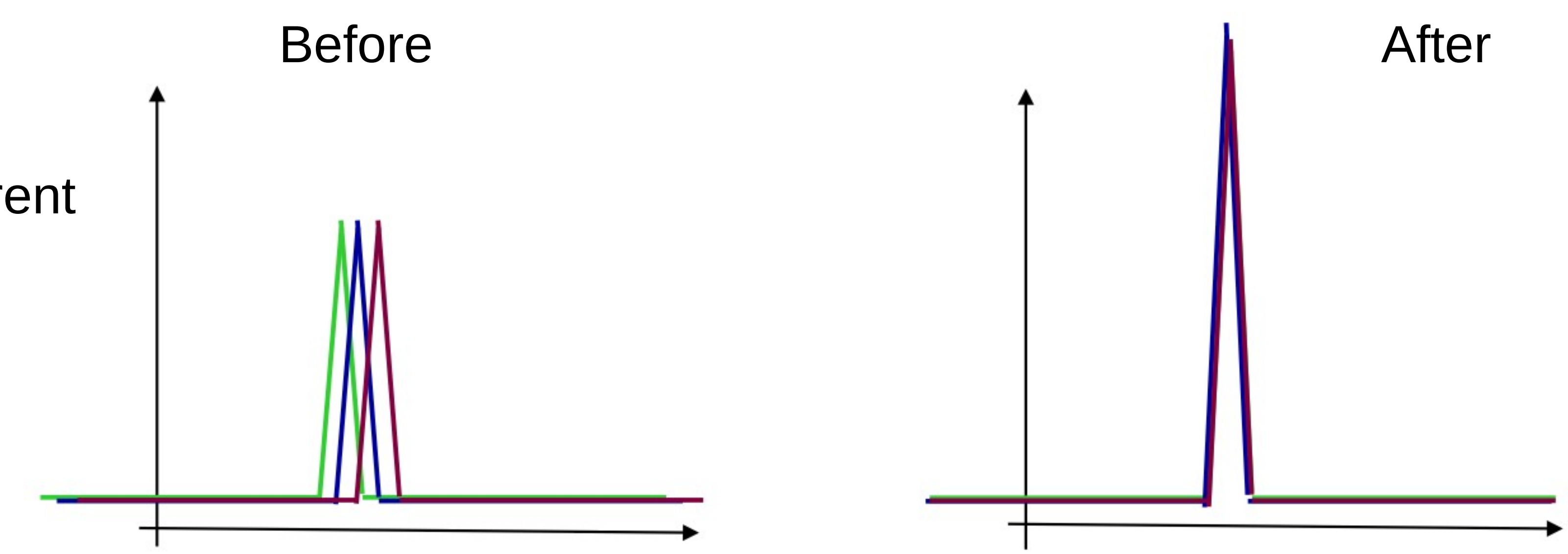
Problems:

- Amplitude different normalisation for different antennas
- Noise levels are high, and are different for different antennas

Doppler correction

- Often now done using **Doppler Setting** and further finer corrections during post-processing and imaging
- Within one observation, Doppler Setting can correct for the Heliocentric velocity (~30 km/s), but cannot exactly correct for the ~0.5 km/s for the Earth’s rotation on its axis
- Remaining corrections done in e.g. tasks clean or cover in CASA and task CVEL in AIPS (after bandpass calibration). CASA task plots can correct on-the-fly (less accurately).

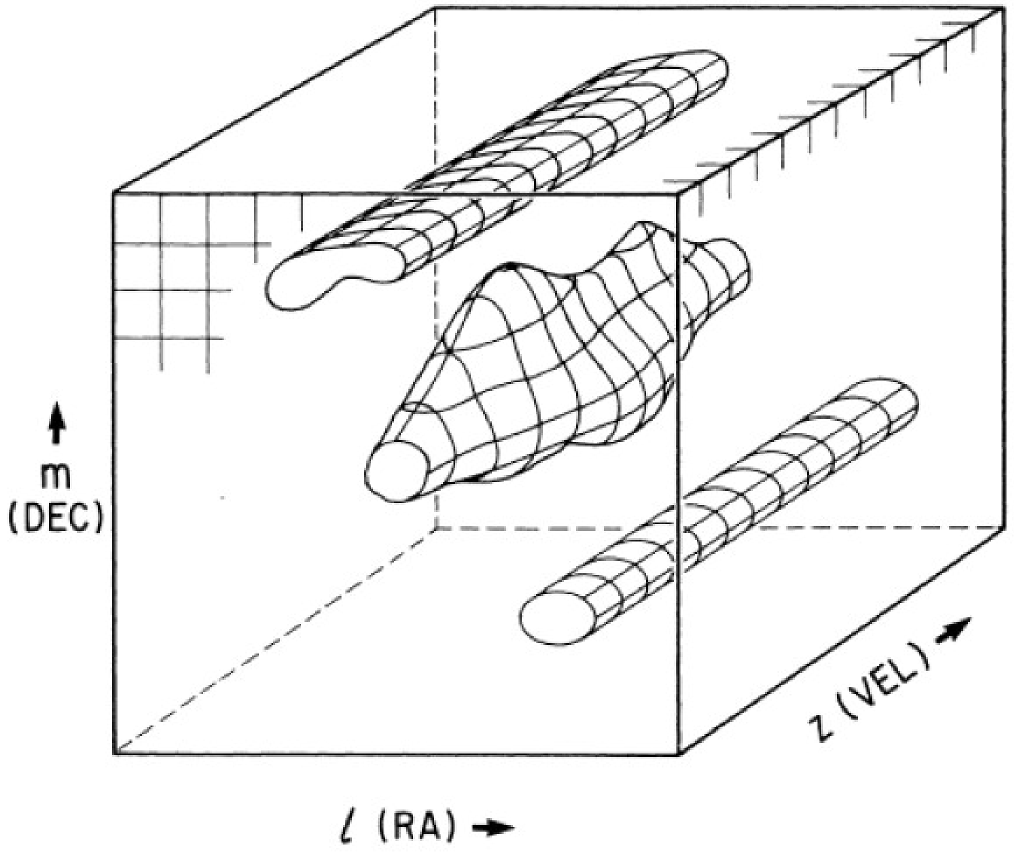
Different colours are different times in the observations



Continuum subtraction

As well as lines, spectral-line data often contains continuum sources (either from the target or from nearby sources in the field)

- This emission complicates the detection and analysis of line data
- Continuum emission limits the achievable spectral dynamic range

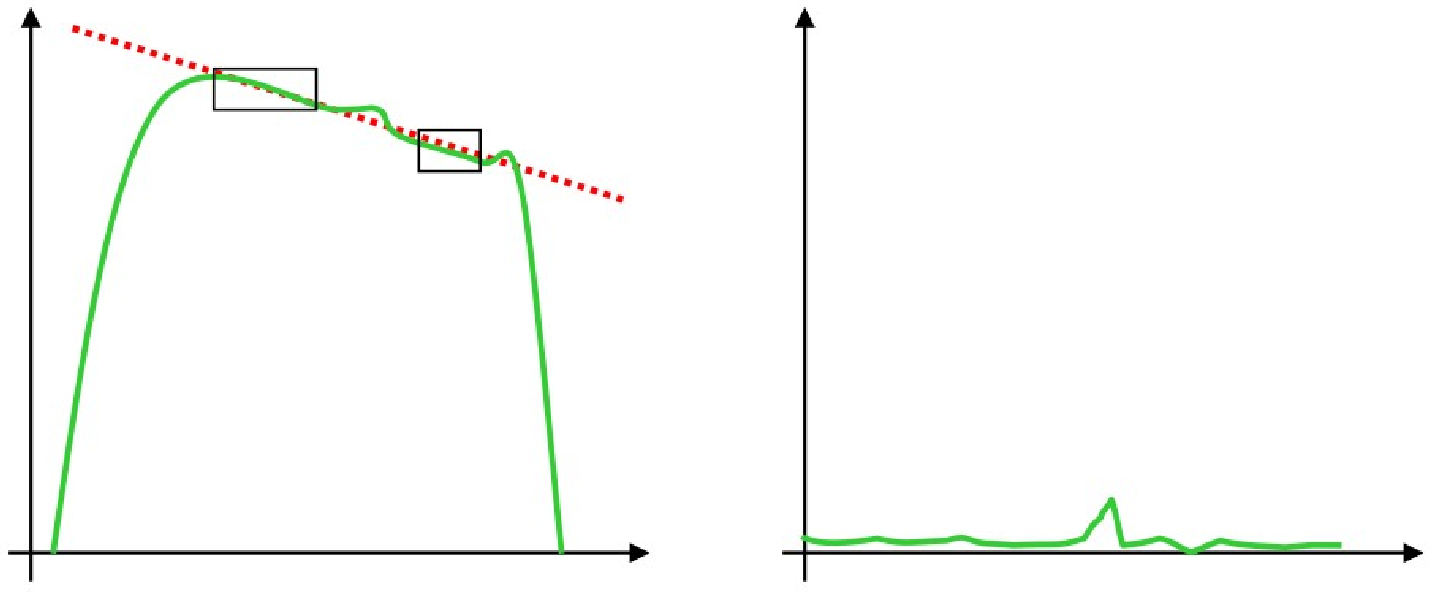


Spectral line cube with two continuum sources (structure independent of frequency) and one spectral line source near the field centre Roelfsma(1989)

Continuum subtraction

Method:

1. examine the data
2. assess which channels appear to be line-free
3. use line-free channels to estimate the continuum level
4. subtract the continuum
5. evaluate the results



Continuum subtraction: Two methods

1. In the *uv*-plane

   subtract continuum then clean line & continuum separately

   - Use AIPS tasks such as UVLIN, UVLSF, UVSUB
   - Use CASA tasks such as uvcontsub

2. In the image-plane

   FT data, subtract continuum from the “dirty” cube then clean both continuum & line

   - Use AIPS task such as IMLIN
   - Use CASA tasks such as imcontsub

No one single subtraction method is appropriate for all experiments!

Self-calibration

Can apply self-calibration solutions to improve quality of line data

Two cases:

1. Strong line emission (i.e. maser)
    - Choose a strong channel with "simple" structure
    - Self-cal that channel & apply solutions to all other channels
    - Allows imaging of weak continuum (& channels) with improved SNR
2. Weak line and strong continuum emission
    - Apply solutions from the continuum to individual channels
    - Allows imaging of weak lines with improved SNR

Get good positions of line features relative to continuum, but lose absolute positional information

Imaging of cubes

- Principles for continuum imaging mostly apply to line data as well (cleaning, weighting, etc.)
- But keep in mind that deconvolution of spectral line data often poses special challenges:
  - Cleaning many channels is computationally expensive (do you need the full spectral resolution or can you average?)
  - Emission *structure* changes from channel to channel (may have to change cleaning boxes for each channel)
  - If you are interested in *both* high sensitivity (to detect faint emission) and high spatial/spectral resolution (to study kinematics):
    - robust weighting with $-1<R<1$ good compromise

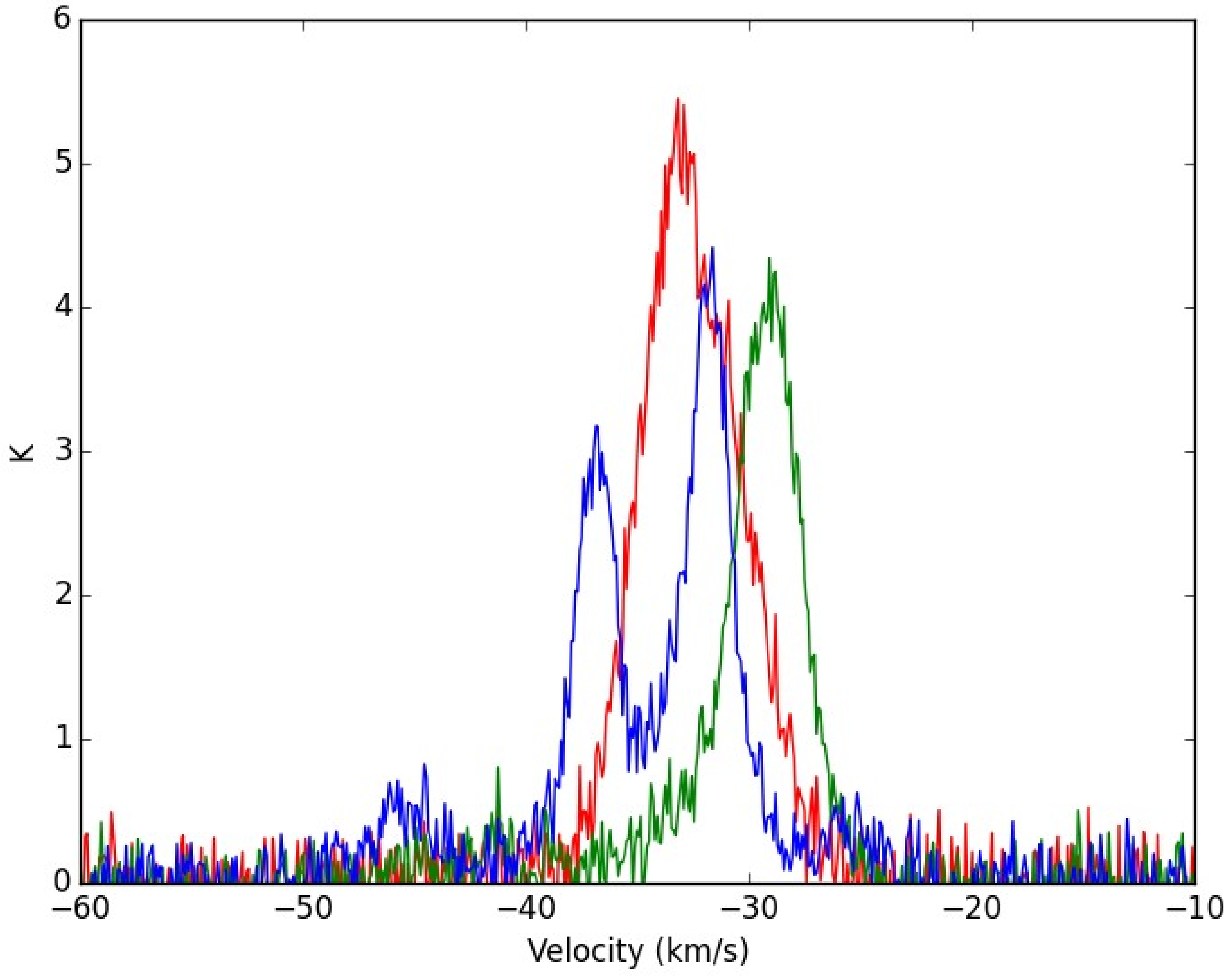
Image and Spectral Analysis

After mapping all channels in the data set, we have a **spectral line 3D data *cube*** (RA, Dec, Vel or Freq)

To visualise the information we usually make 1-D or 2-D projections providing differing analysis methods:

- Line profiles (1-D slices along velocity axis)
- Channel maps (2-D slices along velocity axis)
- Moment maps (integration along the velocity axis)
- Position-velocity (PV) plots (slices along spatial dimension)
- Movies (2-D slices along velocity axis)

Line profiles



- Spectra taken from three different positions in the ^{13}CO (J=1-0) G305 cube.
- Used **KVIS** to output the ascii data for each spectra then made into an image using **Python**

Channel maps

- Channel maps of ^{12}CO (J=1-0) towards G305
- Steps through channels from -57 km/s to -1 km/s
- Integrates emission over 4 km/s

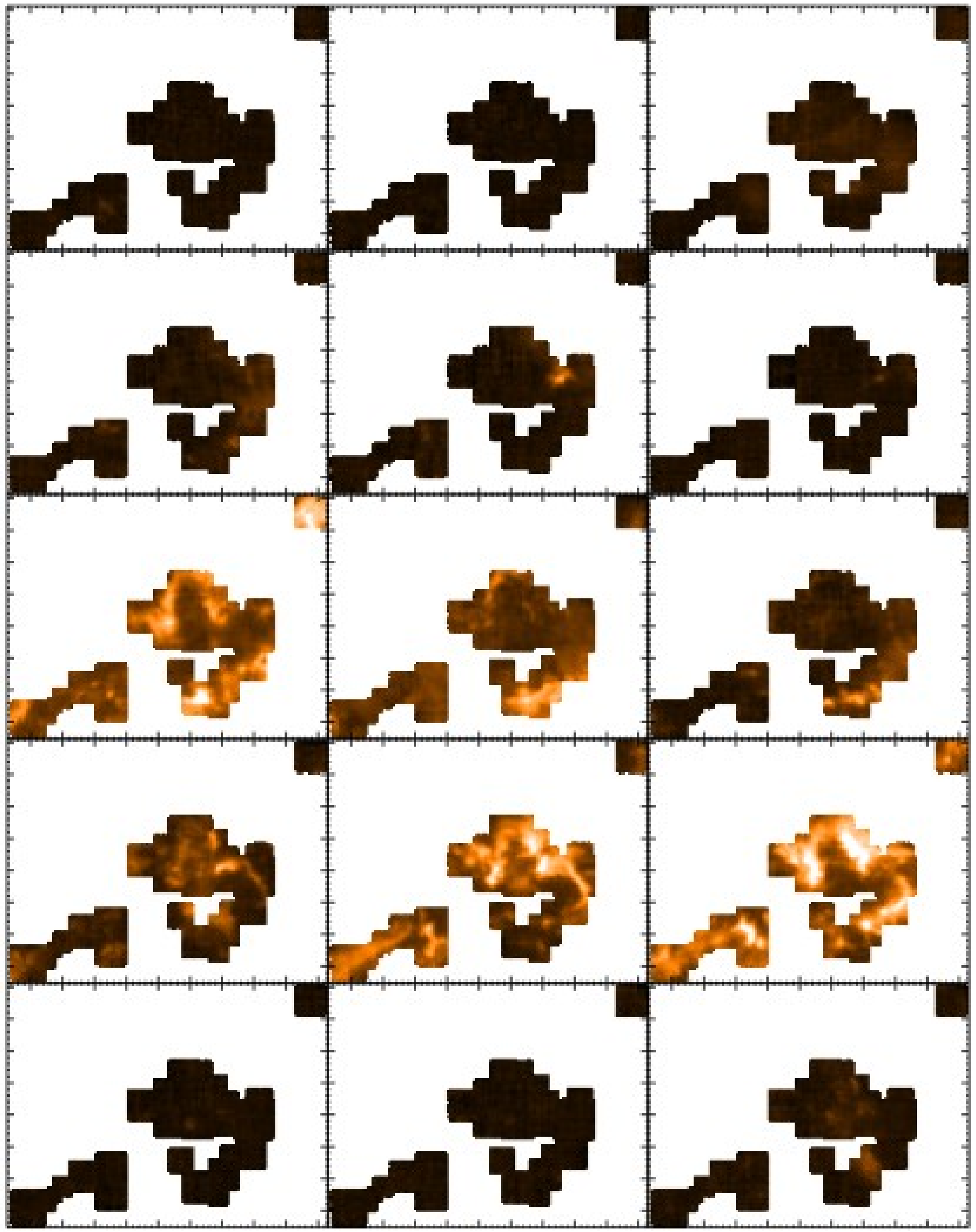


FIGURE 4.13: Channel map of the ^{12}CO data cube beginning at -57.1 km s^{-1} in the bottom left and proceeding to the right and returning to the left at the end of each row. The emission is integrated over 3.7 km s^{-1} channels per image up to -1.2 km s^{-1} in the top right.

Moment maps

Moment maps can be used to derive parameters such as integrated line intensity, centred velocity and line widths as a function of position

$$I_{\rm tot}(\alpha,\delta) = \Delta v \sum_{i=1}^{N_{\rm chan}} S_\nu(\alpha,\delta,\nu_i)$$

Total intensity (moment 0)

$$\overline{v}(\alpha,\delta) = \frac{\sum_{i=1}^{N_{\rm chan}} v_i\, S_\nu(\alpha,\delta,\nu_i)}{\sum_{i=1}^{N_{\rm chan}} S_\nu(\alpha,\delta,\nu_i)}$$

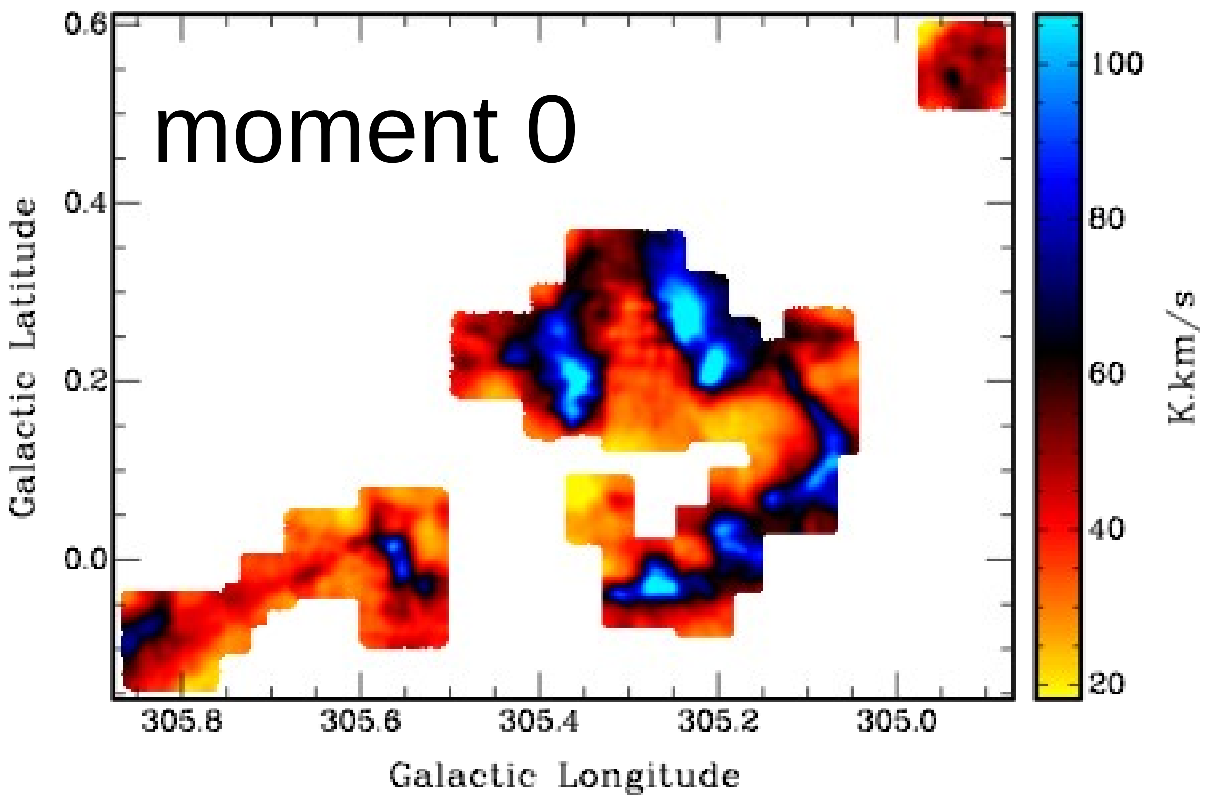
Intensity-weighted velocity (moment 1)

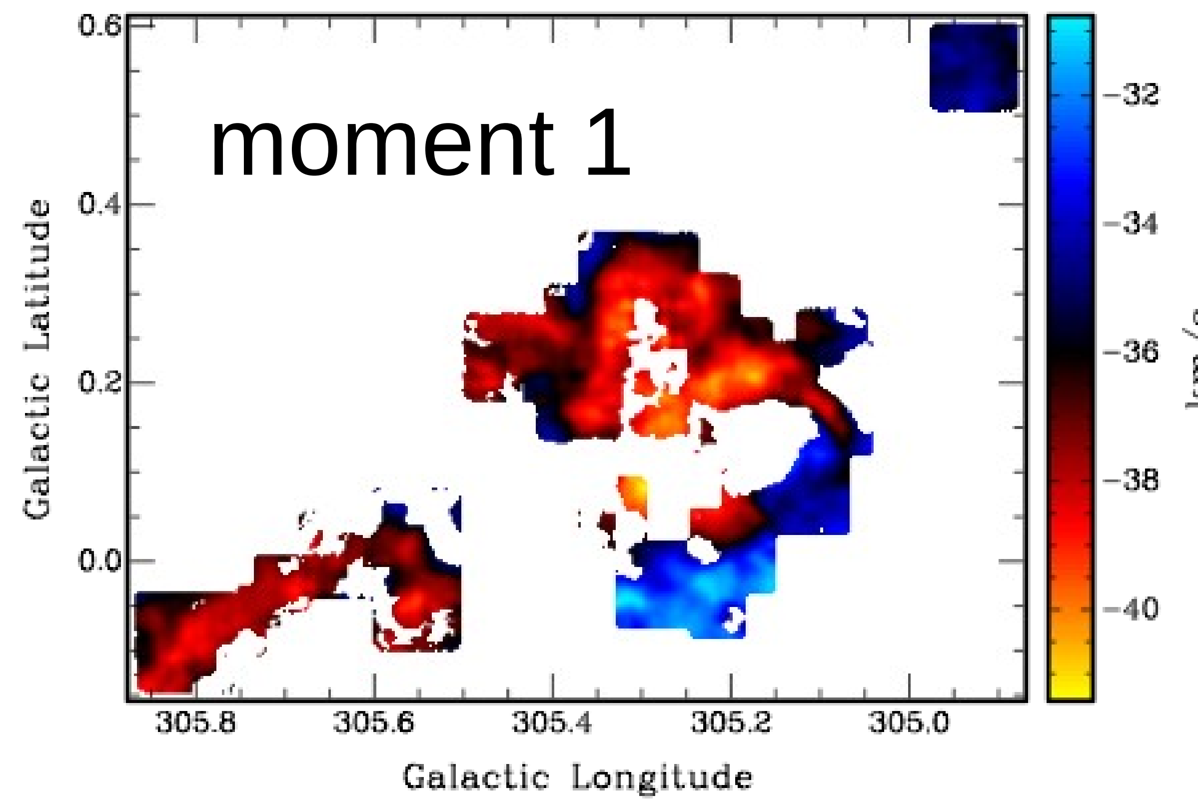
$$\sigma_v(\alpha,\delta) \equiv \sqrt{\langle (v_i - \overline{v}(\alpha,\delta))^2 \rangle} = \sqrt{\frac{\sum_{i=1}^{N_{\rm chan}} (v_i - \overline{v}(\alpha,\delta))^2\ S_\nu(\alpha,\delta,\nu_i)}{\sum_{i=1}^{N_{\rm chan}} S_\nu(\alpha,\delta,\nu_i)}}$$

Intensity-weighted velocity dispersion (moment 2)

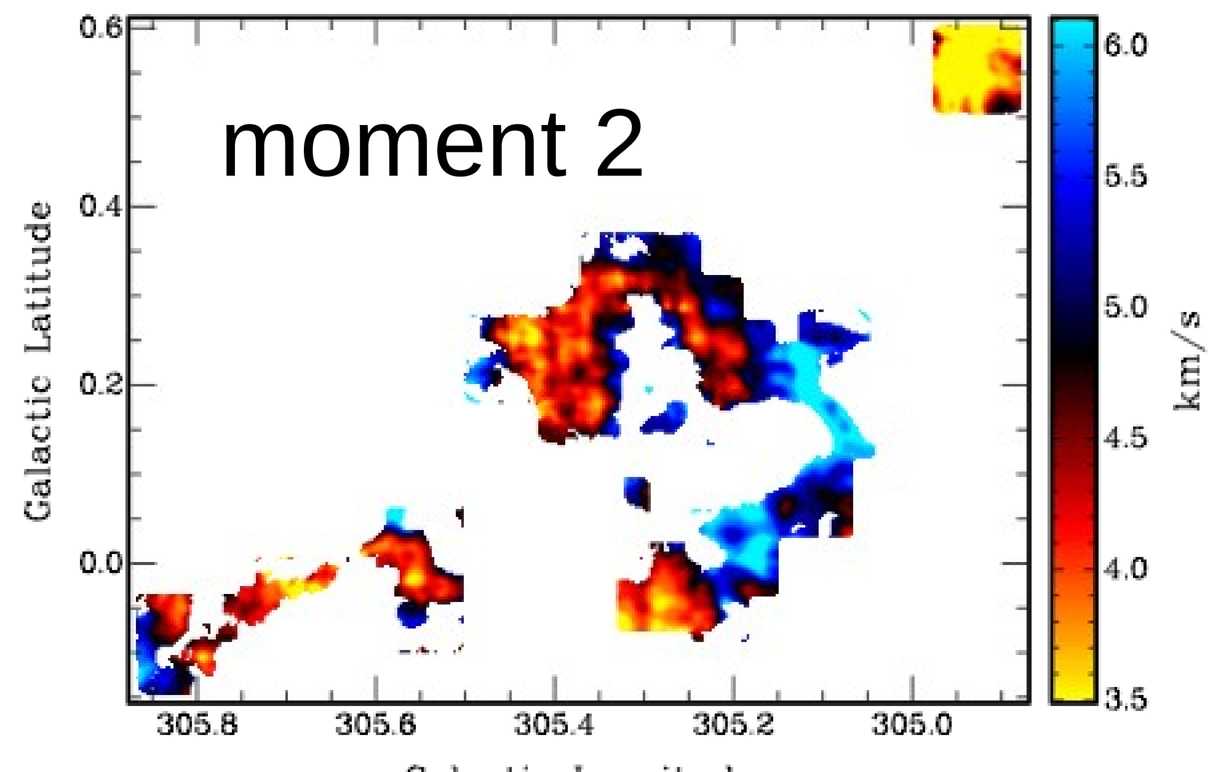
These maps can be made with AIPS tasks like **MOMNT**, CASA's **immoments** or Miriads **moment**

Moment maps





- Moment maps are sensitive to noise so clipping is required
- Sum only over the planes of the data cube that contain emission
- Since higher order moments depend on lower ones (so progressively noise), set a conservative intensity threshold for the 1st and 2nd moments



Position-Velocity (PV) diagrams

- PV diagrams take a slice out of a cube along a **“PV cut”**

- You can produce PV diagrams in CASA using task impv, or IMMOMENTS

- This PV diagram was created in kpvslice

- Try exploring the CO cubes to see if you can spot any interesting velocity structures

    - Edges of molecular clouds?

    - Star forming regions?